Mechanism for the Alteration of the Substrate Specificities of Template-Independent RNA Polymerases

Yukimatsu Toh,1 Daijiro Takeshita,1 Takashi Nagaike,1 Tomoyuki Numata,1 and Kozo Tomita1,2,*
1Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1, Higashi, Tsukuba-shi, Ibaraki 305-8566, Japan
2PRESTO, JST, 4-1-8, Honcho, Kawaguchi-shi, Saitama 332-0012, Japan
*Correspondence: kozo-tomita@aist.go.jp

SUMMARY
PolyA polymerase (PAP) adds a polyA tail onto the 3′-end of RNAs without a nucleic acid template, using adenosine-5′-triphosphate (ATP) as a substrate. The mechanism for the substrate selection by eubacterial PAP remains obscure. Structural and biochemical studies of Escherichia coli PAP (EcPAP) revealed that the shape and size of the nucleobase-interacting pocket of EcPAP are maintained by an intra-molecular hydrogen-network, making it suitable for the accommodation of only ATP, using a single amino acid, Arg197. The pocket structure is sustained by interactions between the catalytic domain and the RNA-binding domain. EcPAP has a flexible basic C-terminal region that contributes to optimal RNA-binding interactions between the catalytic domain and the RNA-binding domain. EcPAP has a flexible basic C-terminal region that contributes to optimal RNA translocation for processive adenosine 5′-triphosphate (ATP) as a substrate. The mechanism for the substrate selection by eubacterial PAP remains obscure. Structural and biochemical studies of Escherichia coli PAP (EcPAP) revealed that the shape and size of the nucleobase-interacting pocket of EcPAP are maintained by an intra-molecular hydrogen-network, making it suitable for the accommodation of only ATP, using a single amino acid, Arg197. The pocket structure is sustained by interactions between the catalytic domain and the RNA-binding domain. EcPAP has a flexible basic C-terminal region that contributes to optimal RNA translocation for processive adenosine 5′-triphosphate (ATP) as a substrate.

INTRODUCTION
Polyadenylation of the 3′ termini of mRNAs by polyA polymerase plays important roles in gene expression in cells. In eukaryotes, the polyadenylation of mRNAs is involved in mRNA stability, maturation, and translation (Beelman and Parker, 1995; Sachs et al., 1997; Wahle and Rüegsegger, 1999). In contrast, in eubacteria, the polyadenylation of mRNAs acts as an RNA-destabilizing signal (Carpousis et al., 1999; Dreyfus and Régnier, 2002; Régnier and Hajnsdorf, 2009). The polyadenylation of RNAs in Escherichia coli is catalyzed by polyA polymerase I, encoded by the pcnB gene (Cao and Sarkar, 1992), and acts as an RNA-destabilizing signal in cells. PolyA polymerase (PAP) is a member of the nucleotidyltransferase super-family, which includes polyA polymerase, CCA-adding enzyme, terminal deoxynucleotidyltransferase (TdT), DNA polymerase β, and kanamycin nucleotidyltransferase (Holm and Sander, 1995; Martin and Keller, 1996, 2004, 2007; Yue et al., 1996). The PAP and CCA-adding enzymes are both template-independent RNA polymerases, and are classified into two classes (classes I and II), according to their active site signatures (Yue et al., 1996). Eukaryotic PAP and archaeal CCA-adding enzymes belong to class I, whereas eubacterial PAP and eubacterial and eukaryotic CCA-adding enzymes belong to class II. Whereas PAP adds the polyA tail onto the 3′-end of any RNAs, using adenosine-5′-triphosphate (ATP) as a substrate, CCA-adding enzyme adds the CCA sequence onto the 3′-end of tRNAs, using cytidine-5′-triphosphate (CTP) and ATP as substrates. Although both the class I and II PAPs catalyze the same reaction, they lack significant similarity in their amino acid sequences.

The crystal structures of the class I PAPs were reported, and revealed the detailed molecular basis for their specificity of nucleotide selection (Bard et al., 2000; Martin et al., 2000; Balbo and Bohm, 2007). However, the crystal structure of a class II PAP has not been solved, and the mechanisms underlying the same activity and specificity, despite the low similarity in the amino acid sequences between the two classes of PAPs, have remained obscure. Compounding the unsolved mechanistic question on the same activity between the divergent class I and II PAPs, the amino acid sequence of the N-terminal half region, about 25 kDa, of the class II PAPs is homologous to that of the class II CCA-adding enzymes (Yue et al., 1996). In the class II CCA-adding enzymes, nucleotide recognition is achieved by Watson-Crick-like hydrogen-bonds between the nucleobases of CTP and ATP and the conserved amino acid residues, Asp and Arg, in the catalytic core pocket (Li et al., 2002; Tomita et al., 2004; Toh et al., 2009). Interestingly, the two key amino acid residues, Asp and Arg, in the catalytic pocket are conserved among the class II RNA-specific nucleotidyltransferases. Therefore, the molecular basis for the different activities between the class II PAP and CCA-adding enzymes also remains obscure.

Here, we present the crystal structures of class II E. coli polyA polymerase (EcPAP) and its complex with ATP. Along with the structure-based mutation studies of EcPAP, we present the molecular basis for the substrate specificity and processivity of class II PAP. A comparison of the structures of class II EcPAP with those of the class II CCA-adding enzymes provides the molecular basis for their different activities. The present study...
also suggests that structural changes in the domain(s) and/or regions outside the well-conserved catalytic domain could change the substrate specificity among the template-independent RNA polymerases.

RESULTS

Structure Determination of EcPAP

The full length EcPAP with the Arg233His mutation was initially crystallized (form I [Arg234His]). Because the Arg234His mutation in EcPAP drastically increased the expression level of the protein in the cells, the mutant EcPAP was crystallized and its structures were determined (see Supplemental Results and Figure S1 available online). The form I (Arg234His) crystal belonged to the space group P6\textsubscript{1}2\textsubscript{2}2. The initial phase was calculated by multiple wavelength anomalous dispersion (MAD) methods, using the selencmethionine EcPAP derivative, at 3.65 Å resolution (Table 1). The experimental electron-density maps corresponding to the N-terminal part of EcPAP were not clear, and it was difficult to trace the C\textsubscript{\alpha} atoms and build the model structure. The crystal with the space group P4\textsubscript{2}12\textsubscript{2} was obtained (form II [Arg234His]) by using the mutant EcPAP lacking the C-terminal 26 amino acids. The form II (Arg234His) crystal yielded X-ray diffraction data up to 2.8 Å resolution. The initial phase calculated using the form I (Arg234His) crystal was applied to the native data set obtained from the crystal form II (Arg234His), by the multicrystal averaging method (Adams et al., 2002). The electron density maps thus obtained, corresponding to the N-terminal part of EcPAP, were easily interpreted. The structure of the form II (Arg234His) crystal was determined at 3.0 Å resolution and refined to an R factor of 23.7% (R\textsubscript{free} = 25.1%). Subsequently, the structure of crystal Arg234 (form II [Arg234]) was also obtained, and its structure conserved three catalytic carboxylates (Asp 69, Asp 71, and Glu 108) (Figure 2A), and was determined at 3.15 Å resolution and refined to an R factor of 27.0% (R\textsubscript{free} = 28.7%) (Table 1).

Overall Structure of EcPAP

The overall structure of EcPAP is composed of head, neck, body, and leg domains, and is referred to as a sea-otter (Figure 1A). The head and neck are the catalytic and nucleobase-interacting domains, respectively, whereas the body and leg domains are the RNA-binding domains. The regions between β\textsubscript{6} and β\textsubscript{4} (amino acid residues 115–137) in the head domains of both apo crystal forms I and II were disordered and remained unresolved. In the apo EcPAP structure from crystal form I, the C-terminal 23 amino acid residues (amino acid residues 432–454) were disordered.

The head domain of EcPAP is composed of several anti-parallel β sheets (β\textsubscript{1}–β\textsubscript{9}) supported by α helices (α\textsubscript{1}–α\textsubscript{12}), and the conserved three catalytic carboxylates (Asp\textsubscript{151}, Asp\textsubscript{271}, and Glu\textsubscript{193}) reside on the β sheets (Figure 2A; Figure S2). The structure of the head domain of EcPAP and the geometry of the three catalytic carboxylates of EcPAP are homologous to those of other nucleotidytransferases (see references in Martin and Keller, 2007). The neck domain of EcPAP is composed of helices (z\textsubscript{2}–z\textsubscript{10}, \textsubscript{13}), and the conserved key amino acid residues (Asp\textsubscript{194} and Arg\textsubscript{197}) among the class II enzymes reside in the neck domain (Figure 2A; Figure S2). The body and leg domains are composed of a bundle of α helices. The C terminus of z\textsubscript{12} in EcPAP protrudes from the body domain, and interacts with the loop between \textsubscript{2} and \textsubscript{10} in the neck domain, as described below. This region in the body domain is referred to as the hand region (Figure 1A). The C-terminal region in the leg domain is disordered and remained unresolved.

The overall structure of EcPAP is fundamentally different from those of the U-shaped class I eukaryotic PAPs (Figure 1B) (Bard et al., 2000; Martin et al., 2000). This observation suggested the divergent evolution of the class I and class II PAPs. The EcPAP structure also differs from those of the U-shaped class I archaeal CCA-adding enzymes (Okabe et al., 2003; Xiong et al., 2003). The overall structure of EcPAP is different from the sea-horse structures of the class II eubacterial and eukaryotic CCA-adding enzymes (Li et al., 2002; Augustin et al., 2003; Tomita et al., 2004; Toh et al., 2009). However, as described below, the head and neck structures of EcPAP resemble those of the class II CCA-adding enzymes (Figure 1B). This reflects the fact that class II CCA-adding enzyme and PAP are homologous in their N-terminal 25 kDa regions (Yue et al., 1996) (Figure S2), although the enzymatic activities of CCA-adding enzymes and PAPs are different.

Catalytic Core Structure of EcPAP

In the structure of the catalytic pocket of apo EcPAP, one magnesium ion was coordinated to the conserved catalytic carboxylates (Asp\textsubscript{151}, Asp\textsubscript{271}, and Glu\textsubscript{193}) (Figure 2A), and was located in almost the same position as that occupied by one of the manganese ions in the structure of the class I PAP (Bard et al., 2000) (Figure S3A). Conventional DNA/RNA polymerization proceeds by the two metal ion catalytic mechanism (Brautigam and Stitz, 1998). The presence of only one magnesium ion in the catalytic site is discussed below. The mutant EcPAPs (Asp\textsubscript{151}Ala, Asp\textsubscript{271}Ala, and Glu\textsubscript{193}Ala) all exhibited reduced adenosine 5′-monophosphate (AMP) incorporation rates into RNAs (Figure 2B). In the structure of the catalytic pocket of EcPAP, the conserved amino acid residues, Asp\textsubscript{151}, in the head domain and Arg\textsubscript{200} in the neck domain, adopt extended conformations, and participate in the formation of an intra-molecular hydrogen-bond network (Figure 2A; Figures S4A and S4B). The N\textsubscript{111} and N\textsubscript{112} atoms of Arg\textsubscript{197} hydrogen-bond with the O\textsubscript{6} and O\textsubscript{2} atoms of Glu\textsubscript{193}, respectively. The N\textsubscript{1} atom of Arg\textsubscript{197} hydrogen-bonds with the O\textsubscript{6} atom of Asp\textsubscript{151}. The O\textsubscript{2} atom of Asp\textsubscript{151} hydrogen-bonds with the N\textsubscript{11} atom of Arg\textsubscript{200}. The N\textsubscript{12} atom of Arg\textsubscript{200} hydrogen-bonds with the O\textsubscript{6} atom of Asp\textsubscript{194}. The structure of the catalytic core pocket of EcPAP is distinct from those of the class II CCA-adding enzymes, as described below. In addition to the hydrogen-bond network in the catalytic core, other intramolecular hydrogen bonds are formed in the neck and head domains (Figure 2A). The N\textsubscript{112} atom of Arg\textsubscript{200} and the O\textsubscript{2} atom of Glu\textsubscript{238} form a hydrogen bond, thus bridging the \textsubscript{2} and \textsubscript{10} helices in the neck domain. The N\textsubscript{112} atom of Arg\textsubscript{194} interacts with the main chain carbonyl oxygen of Arg\textsubscript{113}. The conformation of the C-terminal 23 amino acid residues of EcPAP is different from those of the class II CCA-adding enzymes (Figure 2A). This reflects the fact that class II CCA-adding enzyme and PAP are homologous in their N-terminal 25 kDa regions (Yue et al., 1996) (Figure S2), although the enzymatic activities of CCA-adding enzymes and PAPs are different.
Mechanism for ATP Recognition by EcPAP

ATP was soaked into the apo EcPAP crystal form II (Arg234His), and the structure was analyzed (Figure 2C). The ATP was clearly visible in the α-weighted simulated annealing Fo-Fc omit map of the structure (Figure S4C). In the complex of EcPAP with ATP, one magnesium ion is coordinated by Asp71 and Glu108, and by the α-phosphate of the ATP. The magnesium ion is located at almost the same position as one of the manganese ions that is thought to be involved in the activation of the 3′-OH group of the RNA primer in eukaryotic PAP (Bard et al., 2000) (Figure S3B).

| Table 1. Data Collection, Phasing, and Refinement Statistics for EcPAP Structures |
|-----------------|-----------------|-----------------|
|                 | Form I (Arg234His) | Form II (Arg234His) | Form II (Arg234) |
|                 | SeMet            | PAP I-Apo         | PAPI-Apo         |
| Data collection |                 |                  |                  |
| Space group     | P6_22            | P4_2_2           | P4_2_2          |
| Cell dimensions |                 |                  |                  |
| a = b, c (Å)    | 129.9, 149.7     | 133.0, 176.8     | 133.0, 176.7    |
| α = β, γ (°)    | 90, 120          | 90, 90           | 90, 90          |
| Wavelength (Å)  | 0.97900          | 0.97928          | 0.98317         |
| Resolution (Å)  | 50–3.65 (3.78–3.65) | 50–3.65 (3.78–3.65) | 50–3.65 (3.78–3.65) |
| Rsym            | 0.090 (0.326)    | 0.093 (0.364)    | 0.095 (0.371)   |
| l/σ (f)         | 14.8 (2.0)       | 12.8 (1.8)       | 11.4 (1.6)      |
| Completeness (%)| 98.5 (92.6)      | 97.7 (89.3)      | 96.3 (82.9)     |
| Redundancy      | 4.8 (3.2)        | 4.6 (2.9)        | 4.4 (2.7)       |
| No. reflections | 15,761 (1480)    | 15,647 (1424)    | 15,461 (1327)   |
| Phasing analysis|                 |                  |                  |
| No. of Se sites | 10              |                  |                  |
| Phasing power   |                 |                  |                  |
| Iso (acentric/centric) |       | 0.237/0.259     | 0.673/0.728     |
| Ano             | 1.0174          | 0.968           | 0.123           |
| Rculis          |                 |                  |                  |
| Iso (acentric/centric) |         | 0.850/0.845     | 0.805/0.770     |
| Ano             | 0.770           | 0.944           | 0.996           |
| Mean overall figure of merit (acentric/centric) | 0.358/0.289 |
| Refinement      |                 |                  |                  |
| Resolution (Å)  | 30.0–3.65       | 30.0–3.0         | 30.0–3.3        |
| No. reflections | 8777            | 32,126          | 23,983          |
| Rwork/Rfree     | 27.9/29.5       | 23.7/25.1       | 25.1/27.5       |
| No. atoms       | 3184            | 6344            | 6331            |
| Protein         | 3184            | 6330            | 6330            |
| Nucleotide      | —               | —               | 62              |
| Ion             | —               | 14              | 2               |
| Solvent         | —               | —               | 6               |
| Rmsd            |                 |                  |                  |
| Bond lengths (Å) | 0.011           | 0.009           | 0.008           |
| Bond angles (°) | 1.8             | 1.4             | 1.3             |
| Ramachandran plot|                 |                  |                  |
| Most favored (%)| 83.8            | 89.7            | 88.4            |
| Allowed (%)     | 14.5            | 8.7             | 10.5            |
| Generously allowed (%) | 1.2         | 1.5             | 1.0             |
| Disallowed (%)  | 0.6             | 0.1             | 0.1             |

Rmsd: root-mean-square deviation. Highest resolution shell is shown in parentheses.
As described, in the apo EcPAP structure, only one magnesium ion is coordinated by all three catalytic carboxylates (Asp_{193}, Asp_{194}, and Glu_{197}). On ATP binding to the nucleobase-interacting pocket, the magnesium ion observed in the apo EcPAP structure may move to a position where it could act as a general base to activate the 3'-OH of the RNA primer, as observed in the EcPAP-ATP complex structure. Usually an additional magnesium ion is coordinated to the \( \beta \)- and \( \gamma \)-phosphates of the nucleotide substrate, to stabilize the transition stage of the nucleotidyl transfer reaction. The presence of only one magnesium ion in the present EcPAP-ATP complex structure suggests that the structure represents a stage where ATP has bound to the nucleobase-interacting pocket, but the enzyme is not ready for the nucleotidyl-transfer reaction, due to the absence of an RNA primer. It is likely that RNA polymerization by the class II PAP also proceeds by the conventional two metal ion catalytic mechanism. In the complex structure of EcPAP with ATP, the \( N_1 \) atom of the ATP hydrogen-bonds with the \( N_{111} \) atom of Arg_{197} (distance is 2.60 Å). The adenine base stacks with a hydrogen bond between the Asp_{151} and Arg_{200} residues, thus stabilizing the accommodation of ATP in the pocket. Arg_{197}, Lys_{198}, and Arg_{200} of EcPAP hydrogen-bond with the \( \gamma \)-phosphate group of ATP, and the \( N_6 \) atom of Arg_{150} hydrogen-bonds with the 3'-OH of the ribose of ATP. In the EcPAP complex structure with ATP, the side chains of Asp_{151} and Arg_{200} adopt extended conformations, as also observed in the apo EcPAP structure (Figure 2C), and the \( O_{\delta_1} \) atom of Asp_{194} hydrogen-bonds with the \( N_{112} \) atom of Arg_{200}.

A superposition of the apo and ATP-bound EcPAP structures revealed a conformational change of the side chain of Asp_{194} (Figure 2D). In the ATP-bound EcPAP, the hydrogen-bond between the \( N_{112} \) atom of Arg_{200} and the \( O_{\delta_2} \) atom of Asp_{151} is stacked with the adenine base of ATP. As a result, the carbonyl-group of the side chain of Asp_{194}, which forms hydrogen-bond with the \( N_{112} \) atom of Arg_{200}, rotates by \( \sim 30^\circ \) to accommodate the ATP in the pocket (Figure 2D). In the presence of an RNA primer, the base of the 3'-terminal nucleoside of RNA primer would stack with the adenine base of ATP, thus further stabilizing the ATP accommodation in the pocket (Figure S5).

The mutation of Arg_{197}, which hydrogen-bonds with the \( N_1 \) atom of ATP, to alanine reduced the AMP incorporation rate (Figure 2B). The mutations of Asp_{151} and Arg_{200}, which form a hydrogen bond that stacks with the adenine of ATP, to alanine also exerted the same effect. Moreover, the mutations of Glu_{193} and Asp_{194}, which form an intramolecular hydrogen network, as well as that of the amino acid residue interacting with the 3'-OH of the ribose (Arg_{197}), also reduced the AMP incorporation rate. Because the amino acid residues Glu_{193}, Asp_{194}, and Arg_{197}, in the neck domain of EcPAP, are conserved among the class II enzymes, including the CCA-adding enzyme (Figure S2), we examined the misincorporation of nucleotides other than ATP by EcPAP (Figures 3A and 3B). In the presence of only CTP in the reaction, CMP was incorporated, although the incorporation rate was lower than that of AMP. At a higher concentration of CTP (800 \( \mu M \)), the incorporation rate of CMP was \( \sim 10\% \) of that of AMP. The incorporation rates of uridine-5'-monophosphate or guanosine-5'-monophosphate, in the presence of only the respective nucleotide (uridine-5'-triphosphate [UTP] or guanosine-5'-triphosphate [GTP]), were much lower than that of AMP, even at a high nucleotide concentration (800 \( \mu M \)) (Yehuda-Resheff and Schuster, 2000; Just et al., 2008).

The misincorporations of nucleotides other than ATP by the mutant EcPAP variants (Glut_{193}Ala, Asp_{194}Ala, and Arg_{197}Ala) were also analyzed in the presence of only the respective nucleotide (Figures 3A and 3B). These EcPAP variants misincorporated CTP or UTP significantly in the presence of only the respective nucleotide, and also incorporated GTP, with lower efficiency. However, in the presence of equal concentrations of all four nucleotides (100 \( \mu M \) each), the wild-type EcPAP did not significantly incorporate any nucleotides other than ATP (<5%) (Figure 3C). This reflects the biological significance of the polyaclaylation of RNAs by EcPAP in cells. In contrast, the EcPAP variants significantly misincorporated nucleotides other than ATP in the presence of equal concentrations of all four
nucleotides. CTP was misincorporated up to 15%–45% of the total incorporated nucleotides, in the presence of all four nucleotides. The Asp194Ala and Arg197Ala variants also incorporated UTP, with efficiencies from 5% to 35% of the total incorporated nucleotides. These results suggest that these mutations disrupt the hydrogen-bonding network in the nucleobase-interacting pocket of EcPAP (Figure 2C), and reduce the nucleotide specificity of the EcPAP. Recently, it was reported that the mutations of the conserved residues (Glu193Asn, Asp194Asn, and Arg197Glu) of EcPAP could transform the polyA polymerase to a polyG polymerase in vitro (Cho et al., 2007). These mutations may disrupt the intramolecular hydrogen bond network in the nucleobase-interacting pocket of EcPAP. Therefore, it is likely that the transformation of EcPAP by these mutations might result from the secondary effects of the disruption of the nucleobase-interacting pocket, rather than the direct effect of the new nucleobase recognition by the engineered amino acid residues. Indeed, the EcPAP mutants could incorporate all four nucleotides, although the efficiencies of the respective nucleotide incorporations differed (Cho et al., 2007).

The present complex structure of EcPAP with ATP and the biochemical studies could explain the nucleotide specificity of the class II PAP. Both the size and shape of the nucleobase-interacting pocket discriminate ATP from GTP, using a single amino acid residue, Arg197, in the pocket. GTP is rejected by the enzyme, due to the absence of hydrogen bonds between Arg197 and the nucleobase. The smaller CTP and UTP could be accommodated in the pocket, although they would not snugly fit. As described, CTP or UTP could be misincorporated by EcPAP in the presence of only the respective nucleotide. However, in the presence of all four nucleotides, EcPAP specifically incorporates ATP. Therefore, the complete discrimination of ATP from the other nucleotides by the class II PAP would be achieved by the size and the shape of the nucleobase interacting pocket and the competitive binding of the four nucleotides to the pocket.

RNA Recognition by EcPAP

The electrostatic potential of the EcPAP surface revealed that positively charged areas exist in three regions (regions -I, -II, and -III) on the enzyme (Figure 4A), and would be involved in the RNA binding.

The mutations of Arg203 and Arg234 in region-I to alanine reduced AMP incorporation to <5% of the wild-type level, and the mutation of Glu238 to alanine reduced AMP incorporation to 70% of the wild-type level (Figure 4B). Arg203 hydrogen-bonds with the \( \gamma \)-phosphate of ATP, and also hydrogen-bonds with Glu238 and bridges two helices, \( \alpha_7 \) and \( \alpha_9 \), in the neck domain (Figure 2A). Kinetic analyses of AMP incorporation by the Arg203Ala and Arg234Ala mutants revealed that the AMP incorporation rates increased linearly as the concentration of the RNA primer increased, up to more than five times the \( K_m \) value for

**Figure 2. The Structure of the Catalytic Core Domain of EcPAP and ATP Recognition by EcPAP**

(A) The catalytic core structure of apo EcPAP. The head and neck domains of EcPAP are colored as in Figure 1. The intramolecular hydrogen networks in the nucleobase-interacting pocket are depicted by the dashed lines, with the distances shown in parentheses.

(B) In vitro relative AMP incorporation rates by mutant EcPAP variants. The AMP incorporation rate into the primer RNA by the wild-type EcPAP was defined as 1.0. The bars in the graph are the standard deviations.

(C) ATP recognition by EcPAP. ATP is colored red. The interactions between ATP and the amino acid residues in the catalytic pocket and the intramolecular

**Figure 3. The Structure of EcPAP with ATP and the Biochemical Studies**

(A) The structure of apo EcPAP with ATP. The head and neck domains of EcPAP are colored as in Figure 1. The intramolecular hydrogen networks in the nucleobase-interacting pocket are depicted by the dashed lines, with the distances shown in parentheses.

(B) In vitro relative AMP incorporation rates by mutant EcPAP variants. The AMP incorporation rate into the primer RNA by the wild-type EcPAP was defined as 1.0. The bars in the graph are the standard deviations.

(C) ATP recognition by EcPAP. ATP is colored red. The interactions between ATP and the amino acid residues in the catalytic pocket and the intramolecular

**Figure 4. RNA Recognition by EcPAP**

The electrostatic potential of the EcPAP surface revealed that positively charged areas exist in three regions (regions -I, -II, and -III) on the enzyme (Figure 4A), and would be involved in the RNA binding.

The mutations of Arg203 and Arg234 in region-I to alanine reduced AMP incorporation to <5% of the wild-type level, and the mutation of Glu238 to alanine reduced AMP incorporation to 70% of the wild-type level (Figure 4B). Arg203 hydrogen-bonds with the \( \gamma \)-phosphate of ATP, and also hydrogen-bonds with Glu238 and bridges two helices, \( \alpha_7 \) and \( \alpha_9 \), in the neck domain (Figure 2A). Kinetic analyses of AMP incorporation by the Arg203Ala and Arg234Ala mutants revealed that the AMP incorporation rates increased linearly as the concentration of the RNA primer increased, up to more than five times the \( K_m \) value for

**Figure 5. The Structure of EcPAP with ATP and the Biochemical Studies**

(A) The structure of apo EcPAP with ATP. The head and neck domains of EcPAP are colored as in Figure 1. The intramolecular hydrogen networks in the nucleobase-interacting pocket are depicted by the dashed lines, with the distances shown in parentheses.

(B) In vitro relative AMP incorporation rates by mutant EcPAP variants. The AMP incorporation rate into the primer RNA by the wild-type EcPAP was defined as 1.0. The bars in the graph are the standard deviations.

(C) ATP recognition by EcPAP. ATP is colored red. The interactions between ATP and the amino acid residues in the catalytic pocket and the intramolecular
the RNA primer by the wild-type EcPAP (Figure 4C). The incorporation rates by these EcPAP mutants are drastically lower than that of the wild-type EcPAP, even with a high RNA primer concentration. These results indicate that Arg203 and Arg234 are involved not only in the RNA primer recognition but also in the catalysis. Because these two residues in region-I are proximal to the active site, these two residues would recognize the 3’-region of the phosphate backbone of the RNA primer. These two basic conserved amino acid residues, along with Glu238 in region-I, would be important for relocating the 3’-end of the RNA primer to the catalytic site for RNA polymerization to proceed (Figure S5).

The mutations of Arg300 and Arg354 in region-II to alanine also reduced AMP incorporation to 75% to 50% of the wild-type level (Figure 4B), suggesting that these two residues recognize the phosphate backbone of the RNA. In agreement with this, the kinetic analyses of AMP incorporation by the Arg300Ala and Arg354Ala mutants revealed that the $K_m$ values for the RNA primers by these mutant EcPAPs increased by factors of ~10, relative to the $K_m$ value for the RNA primer by the wild-type (the $K_m$ values for RNA of the Arg300Ala and Arg354Ala EcPAP mutants are ~3.0 and 2.9 $\mu$M, respectively). These results suggest that the Arg300 and Arg354 residues in region-II are involved in the RNA primer binding.

The outer surface of the legs domain, region-III, contains a cluster of positively charged amino acid residues (amino acid residues 368–379) (Figure 4A). The disordered 23 amino acid residues at the C-terminal region of EcPAP are highly positively charged and would be involved in RNA recognition, probably together with the outer surface of the leg domain. The positively charged C-terminal region of EcPAP probably adopts a random-coil structure, as predicted by several programs (McGuffin et al., 2000). We compared the polyadenylated products generated by full-length wild-type EcPAP (Wild-Full) and truncated EcPAP lacking the C-terminal 26 amino acids (Wild-del26aa), using a 23-mer RNA (DN3) as the RNA primer (Figure 4D). Wild-Full incorporates ~600 AMPs at a time. On the other hand, Wild-del26aa exhibited distributed polyadenylation, and three distinct polyadenylated RNA products, ranging from ~42 to 200 nucleotides long, were observed (Figure 4D, upper). A kinetic analysis of polyadenylation by the Wild-Full and Wild-del26aa enzymes revealed that the $K_m$ values for the RNA primer by the truncation of the C-terminal region increased by a factor of two (Figure 4E, upper; $K_m$ values for RNA by Wild-Full and Wild-del26aa are 150 and 320 nM, respectively). This indicates that the C-terminal region is involved in the RNA recognition. On the other hand, the $k_{cat}$ value for AMP incorporation was increased by a factor of around four by the truncation of the C-terminal region. The

Figure 3. Nucleotide Specificity of EcPAP

(A) Misincorporations of nucleotides other than ATP by the wild-type and variant EcPAPs (Glu193Ala, Asp194Ala, and Arg197Ala). The assays were performed in the presence of various nucleotide concentrations (100–800 $\mu$M). The reaction conditions were the same as those in Figure 2B, except for the nucleotide concentration used, and the specific activity of each nucleotide was adjusted to 165 mCi/mmol.

(B) Misincorporation of a nucleotide other than ATP by the wild-type and EcPAP variants in the presence of 100 $\mu$M of the respective nucleotide. The asterisks on the upper gel panel indicate the 32P-labeled nucleotide used. The relative misincorporations were quantified, and the incorporations of the respective nucleotide by the wild-type EcPAP were defined as 1.0 (lower graph).

(C) TLC analysis of the nucleotides incorporated by EcPAP variants (left panels) in the presence of all four 32P-labeled nucleotides and unlabeled nucleotides (100 $\mu$M each). The 32P-labeled products were digested with nuclease P1, and the hydrolyzed products were separated by thin-layer chromatography. The fractions of the nucleotides incorporated into RNA by the EcPAP variants were quantified (right graph). The bars in the graphs are the standard deviations.
Figure 4. RNA Recognition by EcPAP

(A) The electrostatic potential of the surface area of EcPAP (blue, positively charged area; red, negatively charged area). The electrostatic surface model was calculated by the program GRASP2 (Petrey and Honig, 2003). The three positively charged areas (regions -I, -II, and -III) are depicted by circles, and the putative RNA binding channel is shown by a white arrow. A close-up view of the C-terminal domain of EcPAP. The body and leg domains are colored as in Figure 1A. The C-terminal disordered region is enclosed by a dotted circle, and the corresponding amino acid sequence is shown in parentheses.

(B) In vitro relative AMP incorporation rates by EcPAP variants. The AMP incorporation rate into the primer RNA by the wild-type EcPAP was defined as 1.0. The bars in the graph are the standard deviations.

(C) Kinetic analysis of polyadenylation by EcPAP variants (Arg300Ala, Arg354Ala, Arg203Ala, and Arg234Ala). The assays were performed in the presence of various concentrations of RNA primers (0–2.0 μM).

(D) Profiles of polyadenylated products generated by the full-length EcPAP (Wild-Full) and the truncated EcPAP (Wild-del26aa), using DN3 (upper) and oligo A23 (lower) as RNA primers. The assays were performed in the presence of 5.0 μM RNA primer. The graph on the right represents the product profiles analyzed by the BAS-2500 imager (Wild-Full, red; Wild-del26aa, blue). The arrows in the gel panel and the graph indicate the distinct polyadenylated products.

(E) Kinetic analysis of polyadenylation by Wild-Full and Wild-del26aa, using DN3 (upper) and oligo A23 (lower) as RNA primers. The assays were performed in the presence of various concentrations of RNA primers (0–4.0 μM).

See Figures S2, S5, and S6.
same results were obtained when oligoA (A_{23}) was used as the RNA primer (Figures 4D and 4E, lower panels). The increased $k_{\text{cat}}$ by the truncation of the C-terminal region suggests that the 3′-terminal region of an RNA with a longer polyA tail would be compressed on the surface of the full-length enzyme. As a result, the catalysis of the AMP incorporation at the active site might be increased, due to the loss of RNA primer compression by the truncation of the C-terminal region. These results indicate that the C-terminal basic unstructured region of the leg domain in EcPAP could optimize the translocation of the RNA primer for processive AMP-incorporation to the 3′-ends of RNAs (Figure S6).

ATP Recognition by the Class II Enzymes

Although the overall structure of EcPAP is different from that of the class II CCA-adding enzyme (Figures 1A and 1B), the structures of the head and neck domains of EcPAP resemble those of the class II CCA-adding enzyme (Figures 1B and 5A). The superposition of the catalytic core structures of EcPAP and TmCCA revealed striking differences in the conformations of the side chains of two amino acid residues (Asp\textsubscript{151} and Arg\textsubscript{200} in EcPAP and Asp\textsubscript{130} and Arg\textsubscript{180} in TmCCA) (Figure 5B). In the structure of TmCCA, and all other reported structures of CCA-adding enzymes (Li et al., 2002; Augustin et al., 2003; Tomita et al., 2004; Toh et al., 2009), Asp\textsubscript{130} and Arg\textsubscript{180} adopt bent conformations and form a planar hydrogen bond, and Arg\textsubscript{177} and Asp\textsubscript{174} (corresponding to Arg\textsubscript{197} and Asp\textsubscript{194} in EcPAP) do not hydrogen-bond with Asp\textsubscript{130} and Arg\textsubscript{180}, respectively (Figure 5B). On the other hand, in the EcPAP structures (forms I and II), the side chains of the corresponding Asp\textsubscript{151} and Arg\textsubscript{200} residues adopt extended conformations and hydrogen-bond with Arg\textsubscript{197} and Asp\textsubscript{194}, respectively, as described (Figures 2A and 5B).

In the class II CCA-adding enzymes, the 4-NH\textsubscript{2} of CTP and the 6-NH\textsubscript{2} of ATP hydrogen-bond with Asp (Asp\textsubscript{174} in TmCCA), and the N\textsubscript{4} atom of CTP and the N\textsubscript{7} atom of ATP hydrogen-bond with Arg (Arg\textsubscript{177} in TmCCA). The O\textsubscript{2} atom of CTP also hydrogen-bonds with the Arg (Arg\textsubscript{177} in TmCCA). These two residues in the class II CCA-adding enzyme are required for the recognition of both CTP and ATP (Li et al., 2002; Toh et al., 2009). Therefore, the mechanisms for ATP recognition by the class II PAP and CCA-adding enzymes are different. The superposition of the complex structures of EcPAP with ATP and TmCCA with ATP revealed that the tri-phosphate moieties of the ATPs in the complexes superposed well, but the adenine bases did not (Figure 5C). As described above, the side chains of Asp\textsubscript{151} and Arg\textsubscript{200} adopt extended conformations in EcPAP (Figure 2A), and as a result, the adenine base of the ATP is shifted to the catalytic head domain, and the N\textsubscript{4} atom of ATP hydrogen-bonds with the N\textsubscript{7} atom of Arg\textsubscript{197} in EcPAP. In the presence of the RNA primer, the base at the 3′-end of the primer would stack with the adenine base of the ATP, and the ATP would be further stabilized at the ATP incorporation stage, as observed in the class I PAP and class II A-adding enzymes (Balbo and Bohm, 2007; Tomita et al., 2004).

Other Regions in the Head Domain for AMP Incorporation

In the form I and II structures of EcPAP, the region between amino acid residues 114 and 137 in the head domain (between $\beta_6$ and $\alpha_4$) is disordered (Figure 1A; Figure S7A). The corresponding regions in the class II CCA-adding and A-adding enzyme structures are also disordered (Li et al., 2002; Augustin et al., 2003; Tomita et al., 2004; Toh et al., 2009). In the recently solved structure of Thermotoga maritima CCA-adding enzyme, the loop region is clearly visible. Biochemical and genetic analyses demonstrated that the loop region in the catalytic domain is involved in the terminal A\textsubscript{76} addition in vivo and in vitro (Toh et al., 2009). Structural modeling with the RNA primer suggested that the loop region recognizes the 3′-end C\textsubscript{75} and makes the nucleotide-binding pocket suitable only for the accommodation of ATP. Mutations of some of the amino acid residues (Glu\textsubscript{117}, Arg\textsubscript{123}, Arg\textsubscript{128}, and Asp\textsubscript{137}) in the loop region of EcPAP
to alanine did not reduce the AMP incorporation rate significantly (Figure S7B). Thus, the loop region in the catalytic head domain of EcPAP would not be involved in AMP incorporation onto the 3' end of the RNA. These results are consistent with the mechanism for ATP selection by EcPAP, in which both the size and shape of the nucleobase-interacting pocket determine the specificity of nucleotide recognition, in contrast to the mechanism for ATP selection by the class II CCA and A-adding enzymes (Tomita et al., 2004; Toh et al., 2009).

In the class I and class II CCA-adding enzymes, the β-turn in the catalytic domain is involved in all three nucleotide addition steps (Xiong and Steitz, 2004; Cho et al., 2005; Tomita et al., 2006; Martin and Keller, 2007; Toh et al., 2009). The superposition of the EcPAP structure onto that of the *Aquifex aeolicus* A-adding enzyme complexed with tRNA (Tomita and Weiner, 2001; Tomita et al., 2004) revealed that the β-turn is in the vicinity of the 3' end region of the RNA primer (Figure S5). The mutations of Arg93, Arg94, and Phe95, which reside in the β-turn between β5 and β9 in the head domain of EcPAP, to alanine reduced the AMP incorporation rate into RNA to <10% of that of the wild-type enzyme. The Arg93Ala mutation also reduced the AMP incorporation rate to <30% of the wild-type level (Figure S7B). Kinetic analyses of the Arg93Ala and Arg96Ala EcPAP mutants revealed that both mutations increased the $K_m$ values for the RNA primer by >10-fold (the $K_m$ values for RNA of the Arg93Ala and Arg96Ala mutants are 5.7 and 7.9 μM, respectively) (Figure S7C). The Arg96Ala mutation also reduced the $K_m$ by ~4-fold. These results indicate that Arg93 participates in the RNA binding, and Arg96 is involved in both RNA binding and catalysis for AMP incorporation. Therefore, the β-turn in EcPAP is involved in the RNA polymerization and would facilitate the recognition of the 3'-end of the RNA primer during the polymerization reaction, as in the class II CCA-adding enzymes.

**RNA Binding Domains of the Class II Enzymes**

The structure of the body domain of EcPAP is composed of a bundle of α helices (Figure 1A), and is topologically similar to those of several proteins involved in nucleoside metabolism, such as polyphosphate phosphohydrolase (Z-score of 7.2 using the Dali server [Holm and Sander, 1998]), and cGMP (or cAMP)-specific 3', 5'-cyclic phosphodiesterase (Z-score of 6.7). This suggests that the body domain of EcPAP is involved in the recognition of the phosphate group of the nucleoside of RNA.

In contrast to the sequence similarity of the N-terminal half regions (head and neck domains) of the class II PAP and CCA-adding enzymes (Figure S2), that of the C-terminal half of the enzymes is low (Figures 1, 5A, and 6B). However, the structures of the body domains of EcPAP and CCA-adding enzyme share similar topologies and folds (Figures 6A and 6B). The helix lengths and their arrangements in the body domains are different between EcPAP and CCA-adding enzyme (Figures 6A and 6B). One of the prominent differences between the structures in the body domains is found in the helix α12 (in EcPAP). The C-terminal half of α12 and the loop between α12 and α13 of EcPAP are referred to as the hand region. In the structure of the EcPAP body domain, helix α12 adopts a straight conformation (Figure 6A, left), and the O$_{\text{D2}}$ atom of Asp292 and the main chain carbonyl oxygen of Val301 in the hand region interact with the main chain amide of Gly244 and the N$_{\text{C2}}$ atom of Gln244 in the neck domain, respectively. Moreover, His297 in the hand region stacks with Tyr247 in the loop between α9 and α10, and Tyr247 interacts with the side chain of Lys207 in α7 (Figure 6A, right). On the other
hand, the corresponding helix α14 in TmCCA is curved inward, and interacts with α19 and α15 (Figure 6B).

An in vitro biochemical study, using chimeric enzymes of E. coli CCA-adding enzyme and PAP, revealed that helix M (corresponding to α14 in TmCCA and α12 in EcPAP) in the body domain could dictate whether the enzyme synthesizes the polyA or CCA sequence (Betat et al., 2004). It was proposed that, whereas α14 and/or the surrounding regions in the CCA-adding enzyme might trigger the rotation of the side chains of Asp194 and Arg197, the α12 and/or surrounding regions in EcPAP would not. The mutations of Tyr247 and His296 to alanine did not significantly reduce the AMP incorporation rate, and the deletions of the amino acid residues in the hand region reduced the AMP incorporation rates to 10% of the wild-type level (Figure 6C). However, these variants did not reduce the nucleotide specificity (Figure 6D). These results suggest that the interaction of helix α12 with α10 in EcPAP alone does not dictate the specificity of EcPAP for only ATP. Together with the recent results, obtained using EcPAP and CCA-adding enzyme chimeras, it is likely that α12, in the context of other helices in the body domain of EcPAP, collaboratively dictates the nucleotide specificity of EcPAP.

In the body structure of EcPAP, α12 and α14, followed by α15, bundle with α16 and α18 (Figure 6A). On the other hand, the corresponding helices in CCA-adding enzyme, α15 and α16, are shorter (Figure 6B), and do not bundle with α17, as observed in EcPAP. Together, the tight assembly of the α helices following helix α12 in the body domain in EcPAP, and the interaction of α12 with α10 in the neck domain, might prevent the rotation of the side chains of Asp194 and Arg197, and maintain the structure of the nucleobase-interacting pocket by unknown mechanisms. Clarification of the mechanism for maintaining the structure of the nucleobase-interacting pocket suitable for only ATP will await the structure of class II PAP with an RNA primer and an incoming ATP.

Another difference in the C-terminal half regions of the class II PAP and CCA-adding enzymes exists in the C-terminal regions (Figures 1, 5A, and 6). As described, in EcPAP, the C-terminal region of the leg domain seems to be unstructured, and is involved in the optimization of RNA translocation for the processive polyadenylation of RNAs (Figure 4). On the other hand, the C-terminal region of TmCCA (the tail domain) is structured, and is composed of α helices (Figure 6B). The tail domain of the CCA-adding enzyme precisely interacts with the TψC loop of tRNA, functioning as an anchor to prevent the RNA from dislodging from the enzyme surface during CCA addition, and the body domain interacts with the TψC-acceptor helix of tRNAs (Tomita et al., 2004; Cho et al., 2006). These structural variations form the basis for the differences in the primer specificity and the processivity of the class II enzymes.

DISCUSSION
The present structure of class II eubacterial PAP highlights the molecular basis for the different substrate specificities between the class II PAP and CCA-adding enzymes (Figure 7). The nucleobase-interacting pocket of EcPAP is composed of an intramolecular hydrogen bond network, and both the size and shape of the pocket are restricted for the preferential accommodation of ATP. The conserved Arg197 in the pocket recognizes the N1 atom of ATP, and Asp194 participates as a structural element in the scaffold of the nucleobase-interacting pocket, rather than as a recognition element for the 6-NH₂ group of ATP. The mechanism for nucleotide selection by EcPAP is different from that of the eubacterial CCA-adding enzyme. In the class II CCA-adding enzyme, the pocket structure allows the enzyme to incorporate both CTP and AMP, using both Asp and Arg in the pocket. The interaction between the body and neck domains of EcPAP, mediated by the hand region, may contribute to sustaining the structure of the nucleobase-interacting pocket for the accommodation of only ATP. The absence of these interactions in the CCA-adding enzyme might increase the flexibility of the nucleobase-interacting pocket of the CCA-adding enzyme so it can accommodate both CTP and ATP, depending on the reaction stage. The C-terminal region of EcPAP would optimize RNA translocation for the processive polyadenylation of RNAs. The corresponding C-terminal region of the class II CCA-adding enzyme forms α helices, and acts as an anchor that interacts with the TψC loop of tRNA and prevents the RNAs from dislodging from the enzyme surface during RNA polymerization. The absence of the anchoring domain and the presence of the unstructured C-terminal region, which potentially contributes to the optimal translocation of RNA, would allow EcPAP to add multiple AMPs to the 3’-end of any RNA primer. The present results also suggest that during the evolutionary process, structural changes in the protein domain(s) and/or regions outside the well-conserved catalytic domain might have altered and/or restricted the substrate specificities of the template-independent RNA polymerases.
EXPERIMENTAL PROCEDURES

Preparation of EcPAP

The E. coli pcnB gene, encoding polyA polymerase I (EcPAP), was PCR-amplified from genomic DNA and cloned into the pET15b vector between the Ndel and Xhol sites. Because E. coli PAP I reportedly undergoes proteolytic maturation (Cao and Sarkar, 1992), the second valine residue (Val12) from the N terminus was mutated to the initiation Met, and the residues were renumbered. The expression level of the full-length, wild-type PAP was low. During the PCR amplification, cloning and expression of EcPAP, we isolated a clone with a single amino acid mutation (Arg234His). The expression level of EcPAP with the Arg234His mutation was much higher than that of wild-type EcPAP. Therefore, for the initial crystallization and structure determination of PAP, we used EcPAP (Arg234His) in this study (Figure S1). Afterward, we also purified and crystallized wild-type EcPAP. EcPAP was expressed, purified by several types of column chromatography (Supplemental Results), and stored at –80°C until use.

Crystallization and Data Collection

For the crystallization of EcPAP, 1 μl of protein solution (5 mg/ml) was mixed with 1 μl of reservoir solution, containing 100 mM Tris-Cl, pH 8.4, 0.5 M sodium chloride, 5 mM MgCl₂, and 10% (w/v) polyethylene glycol, and the drop solution was equilibrated against the reservoir solution at 20°C by the hanging drop vapor diffusion method (form I Arg234His crystal), for >3 months. EcPAP lacks a nitrogen stream, and the data were collected at the beam-lines BL-17A of Structure Determination of EcPAP 2.8 Å resolution, by multicrystal averaging with PHENIX (Adams et al., 2002). The model was refined by the program CNS (Brunger et al., 1998). We tried to obtain the complex crystal of EcPAP with ATP. However, we could not grow the complex crystals under the same crystallization conditions. Therefore, we soaked ATP into the apo EcPAP crystal. The refinement statistics are shown in Table 1. The structure of form II Arg234 was also determined, and was refined as described.

ACCESSION NUMBERS

The atomic coordinates and structural factors have been deposited in the Protein Data Bank, under the accession codes 3AQK, 3AQL, 3AQM, and 3AQH.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Results and seven figures and can be found with this article online at doi:10.1016/j.str.2010.12.006.

ACKNOWLEDGMENTS

We thank Azusa Hamada of AIST for technical assistance. We thank the beam-line staffs of BL-17A (KEK, Tsukuba) for technical assistance during data collection. This work was supported by grants from the PRESTO program of JST, JSPS, Toray Science Foundation, Takeda Science Foundation and Mochida Memorial Foundation for Medical and Pharmaceutical Research to K.T.

REFERENCES


Yue, D., Maizels, N., and Weiner, A.M. (1996). CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the CCA-adding enzyme from the archaeal hyperthermophile Sulfolobus shibatae. RNA 2, 895–908.